The Pumping Lemma

Mark Greenstreet, CpSc 421, Term 1, 2008/09

Lecture Outline

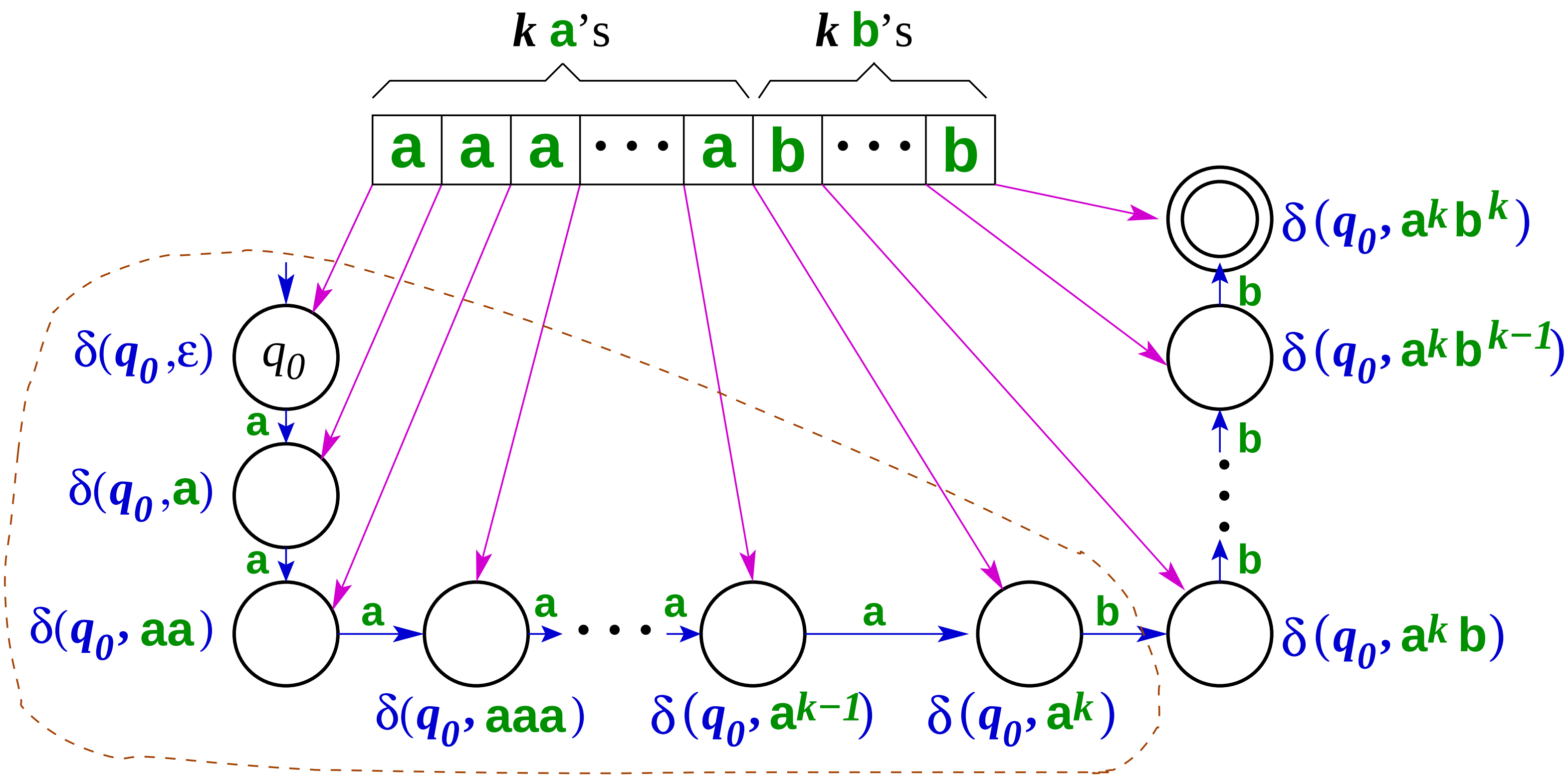
Regular Expressions

- An Example
- The Pumping Lemma

$\mathsf{a}^n\mathsf{b}^n$ is not regular

- Let $\Sigma = \{\mathsf{a}, \mathsf{b}\}$.
- Let $A = \{s \in \Sigma^* \mid \exists n \in \mathbb{N}.\ s = \mathsf{a}^n\mathsf{b}^n\}$.
- At the end of the Sept. 19 lecture, we showed that A is not regular.
- We'll repeat that proof in more detail and show how it can be generalized to give a very powerful way to show that a language is not regular.
  - If A is regular then there is some DFA, M that recognizes A.
  - M has some fixed number of states. Let k be this number.
  - Consider what happens when M reads $\mathsf{a}^k\mathsf{b}^k$.

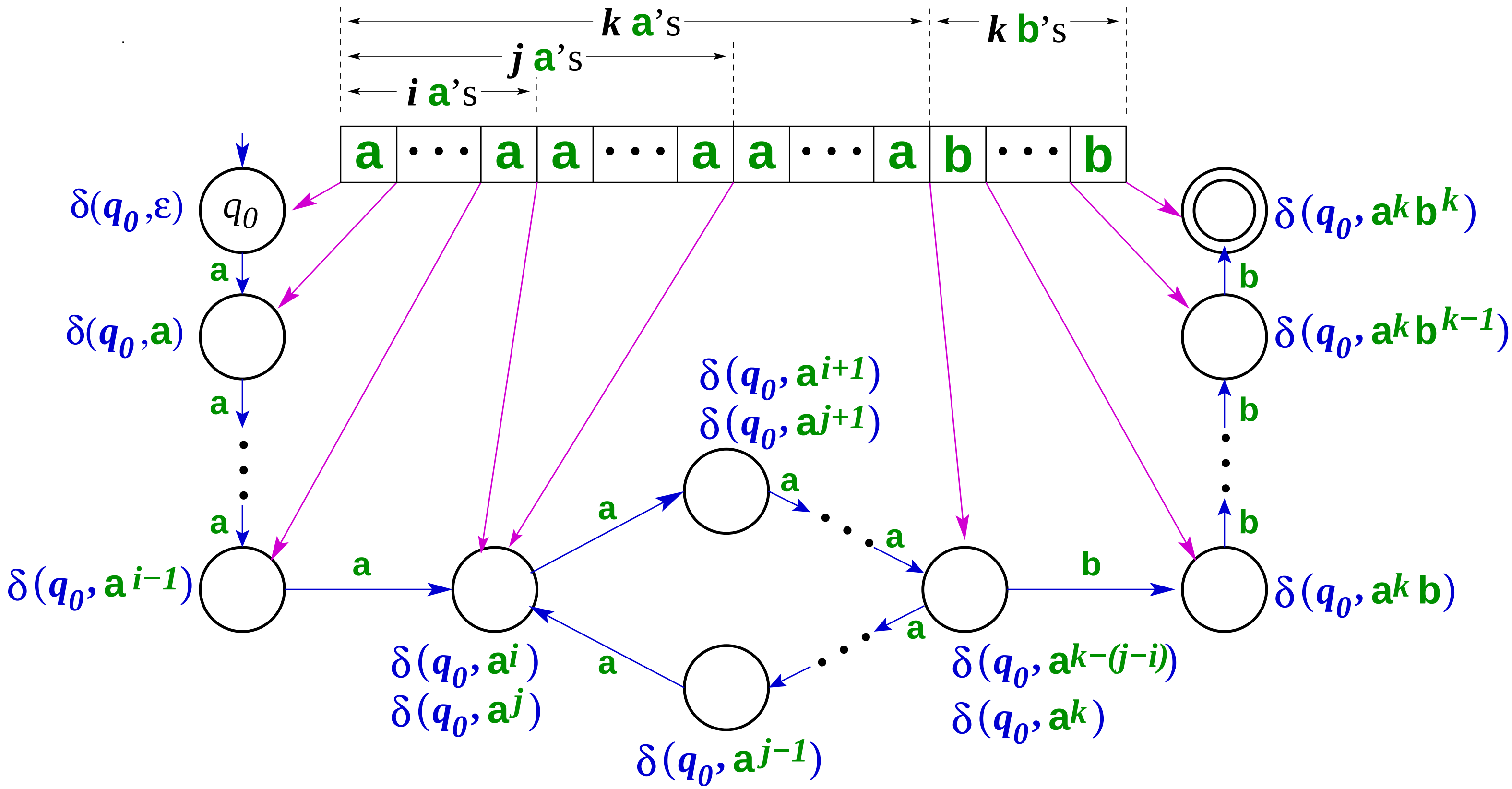
Reading $a^n b^n$



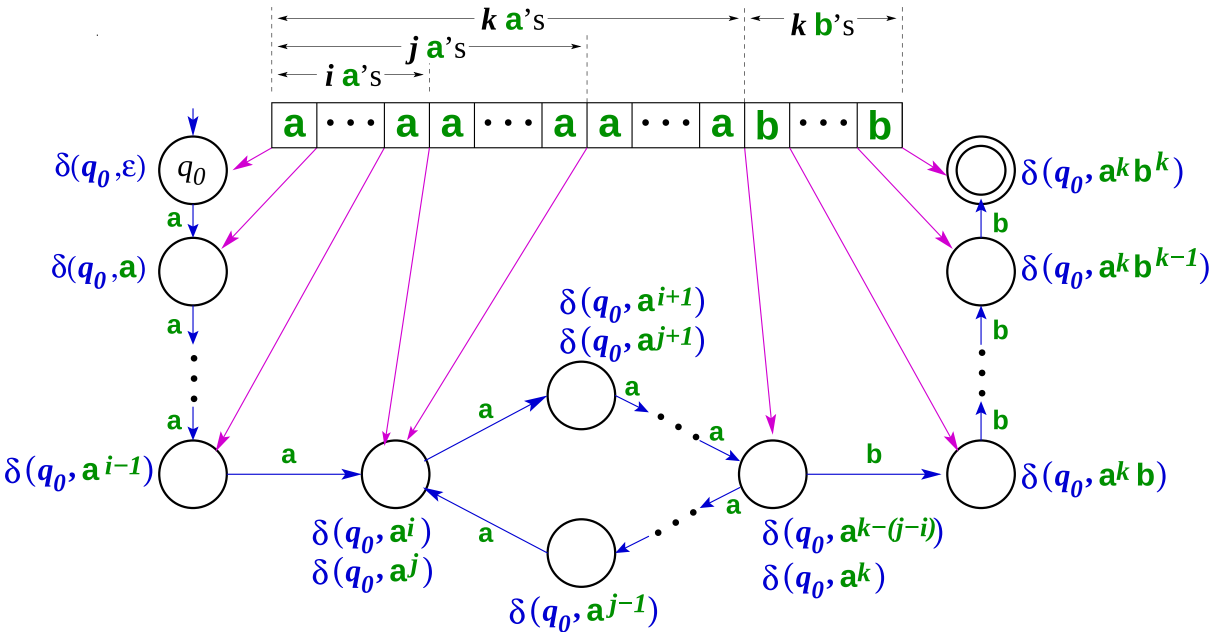
$k+1$ states: they can't all be different!

By the "Pigeon Hole" principle, there must be i and j with $0 \leq i, j \leq k$ and $i < j$ such that the DFA reaches the same state after reading a^i and after reading a^j.

Reading $a^n b^n$



Reading $a^n b^n$



- Let $d = j - i$. We chose $i < j$; therefore $d > 0$.
- If M accepts $a^k b^k$, then
  - M also accepts $a^{k-d} b^k$.
  - M also accepts $a^{k+hd} b^k$, for any $h \geq 0$.
- Thus, DFA M does not recognize language A.
- Therefore, A is not regular.

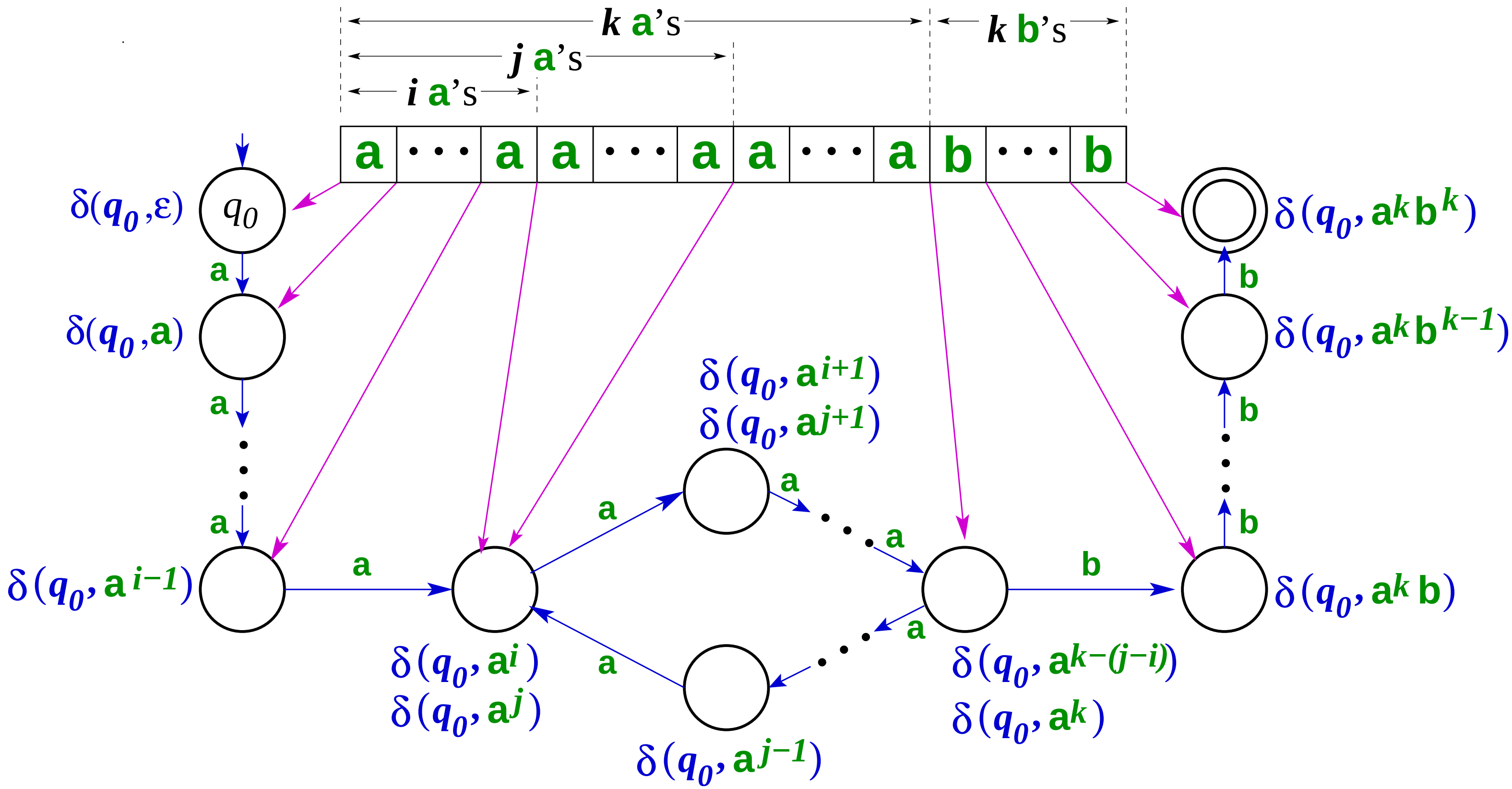
Another non-regular language

- Let $\Sigma = \{a, b\}$ as before.
- Let $B = \{s \in \Sigma^* \;\; \#a(s) = \#b(s)\}$.
  ($\#a(s)$ denotes the number of a's in s and likewise for $\#b(s)$.)
- B is not regular.
- Proof:
  - $B \cap (a^*b^*) \; = \; a^n b^n$.
  - a^*b^* is regular.
  - The regular languages are closed under intersection.
  - If B were regular, $a^n b^n$ would be regular as well.
  - $a^n b^n$ is not regular.
  - Therefore, B is not regular.

How about $w \cdot w^{\mathcal{R}}$?

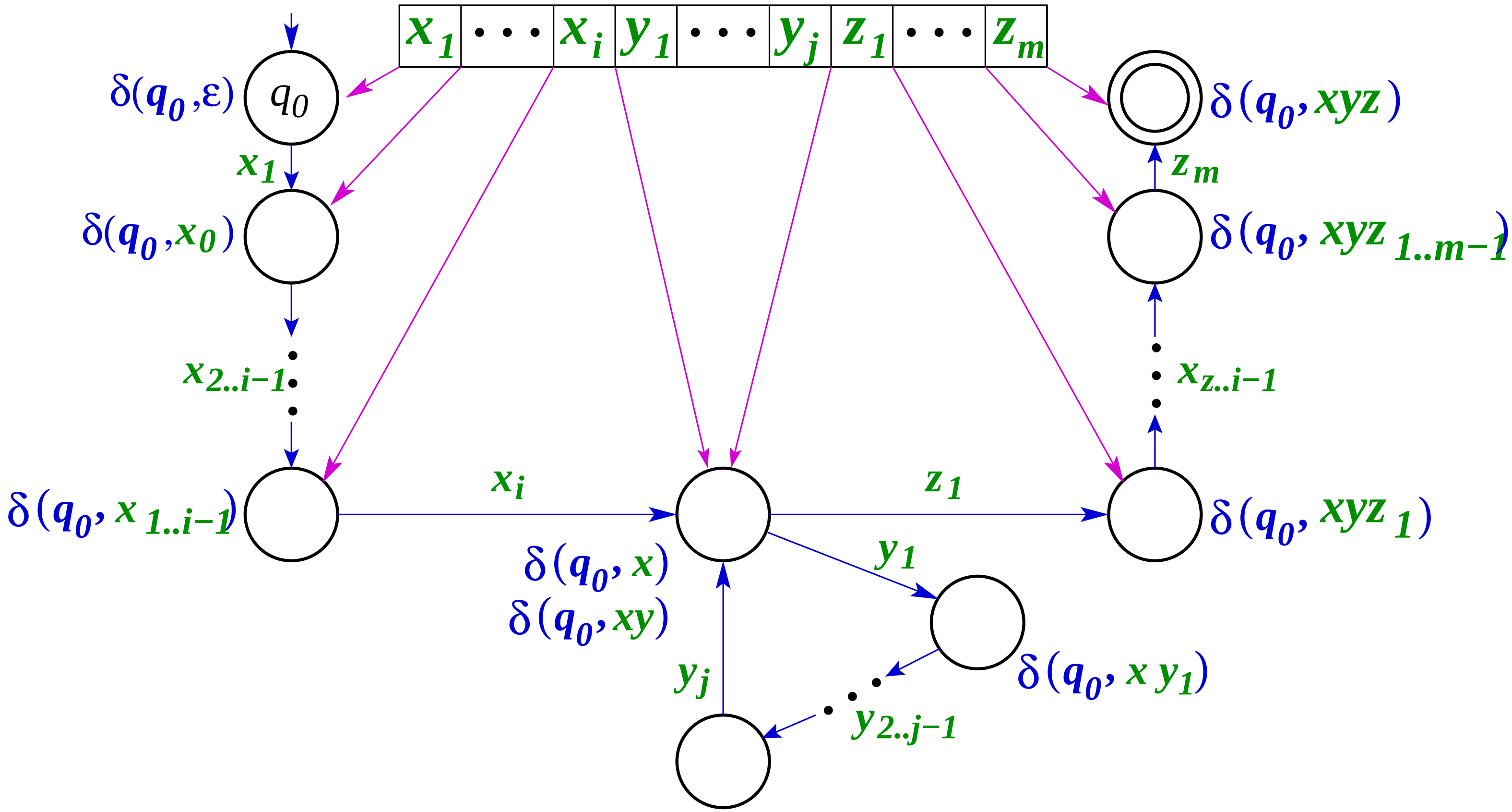
- Let $\Sigma = \{\text{a}, \text{b}\}$ as before.
- Let $C = \{s \in \Sigma^* \;\; \exists w \in \Sigma.\ s = w \cdot w^{\mathcal{R}}\}$.
- C is not regular, but there isn't an obvious way to conclude from knowing that a^nb^n is not regular.
- Instead, we can generalize our earlier proof.
- The generalization is called the pumping lemma and it gives us a very flexible tool for showing that languages are not regular.

A more general argument



A more general argument

$$w = xyz, \quad |w| \geq k$$



The Pumping Lemma

Let A be a regular language.

- There exists some integer p such that for any string w in language A with $|w| \geq p \ldots$
- ...we can find strings x, y, and z such that $w = xyz$ and
  - $\forall i \geq 0.\ xy^i z \in A$,
  - $|y| > 0$, and
  - $|xy| \leq p$.
- The intuition behind the pumping lemma is that:
  - y is a string that takes a DFA that recognizes A through a cycle of states (i.e. a loop).
  - If $|w|$ is greater than the number of states in a DFA that recognizes A, then the DFA must visit the some state more than once when reading w.

    This provides the cycle.

Pumping Lemma: Some Definitions

- Given a regular language A,
- let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA that recognizes A.
- Let $p = |Q|$.
- Let w be any string in A with $|w| \geq p$.
- Let

$$\begin{array}{rcll} \mathit{prefix}(w,n) & = & w, & \text{if } |w| \leq n \\ & = & \mathit{prefix}(x,n), & \text{if } w = x \cdot c \text{ and } |w| > n \\ \mathit{suffix}(w,n) & = & \multicolumn{2}{l}{\text{the string } s \text{ such that } w = \mathit{prefix}(w,n) \cdot s} \\ \mathit{substring}(w,i,j) & = & \mathit{suffix}(\mathit{prefix}(w,j),i) & \end{array}$$

Informally (ignoring special cases for short strings handled above),

$$\begin{array}{rcl} \mathit{prefix}(w,n) & = & w_0 \cdot w_2 \cdots w_{n-1} \\ \mathit{suffix}(w,n) & = & w_n \cdot \cdots w_{|w|-1} \\ \mathit{substring}(w,i,j) & = & w_i \cdot \cdots w_{j-1} \end{array}$$

Pumping Lemma: The Proof

- Note that M must visit some state twice by the time it has read $\mathit{prefix}(w, p)$. This is because M only has p states, and it has visited $p + 1$ states (including the start state) by the time it reads $\mathit{prefix}(w, p)$.
- Let $0 \le i < j \le p$ be integers such that

$$\delta(q_0, \mathit{prefix}(w, i)) \quad = \quad \delta(q_0, \mathit{prefix}(w, i))$$

- Now, let

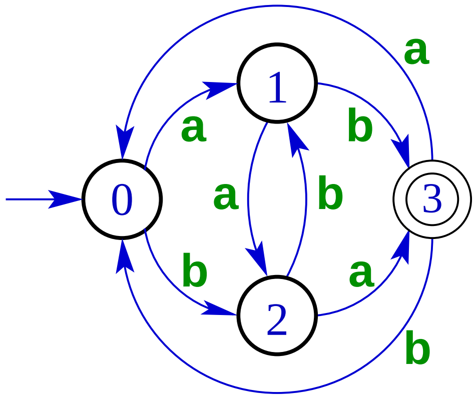
$$\begin{aligned} x &= \mathit{prefix}(w, i) \\ y &= \mathit{substring}(w, i, j) \\ z &= \mathit{suffix}(w, j) \end{aligned}$$

- We have
  - $w = xyz$: by the definitions of x, y, and z.
  - $xy^k z \in A$: see the next slide.
  - $|y| > 0$: $i < j$.
  - $|xy| \le p$: $j \le p$.
- Thus, the claims of the pumping lemma are satisfied.

Proof that $xy^k z \in A$

- Let $q_i = \delta(q_0, x)$.
- $q_i = \delta(q_0, xy) = \delta(\delta(q_0, x), y) = \delta(q_i, y)$.
  In short, $\delta(q_i, y) = q_i$.
- $\delta(q_0, xy^i) = q_i$, by induction on k:
  - Base case, $k = 0$: $\delta(q_0, xy^0) = \delta(q_0, x) = q_i$.
  - Induction step: $k > 0$:
    $\delta(q_0, xy^k) = \delta(q_0, xy^{k-1}y) = \delta(\delta(q_0, xy^{k-1}), y) = \delta(q_i, y) = q_i$
- $\delta(q_0, xy^i z) = \delta(q_i, z) = \delta(\delta(q_0, xy), z) \in A$.
- $\square$

Pumping Lemma: Example



- Let $M = (Q, \Sigma, \delta, q_0, F)$ be the DFA shown above, with $Q = \{0, 1, 2, 3\}$, $\Sigma = \{$a, b$\}$, $q_0 = 0$, and $F = \{3\}$.
- Let $A = L(M)$. Let $p = |Q| = 4$.
- Let $w =$ aabaa. Note that $w \in A$.
- We can show that the claims of the pumping lemma are satisfied by choosing $x =$ a, $y =$ ab and $z =$ aa.
  - $\forall i.\ xy^iz \in A$.
  - $|y| = 2 > 0$.
  - $|xy| = 3 < 4 = p$.

Using the Pumping Lemma

- Typically, we use the pumping lemma to show that a language is not regular.
- To do so, we use the contrapositive of the pumping lemma:
  - If it is not possible to choose an integer p such that for any string $w \in A$ there are strings x, y, z such that
    - $w = xyz$,
    - $\forall i.\ xy^i z \in A$,
    - $|y| > 0$, and
    - $|xy| \leq p$.
  - then A is not a regular language.
  - Note that p is chosen first, and then w can be chosen according to the choice of p.
  - Typically, we find a w (depending on the choice of p) such that there is no way to break w into x, y, and z such that $\forall i.\ xy^i z \in A$.
  - Often, the counterexample uses $i = 2$ or $i = 0$.

$w \cdot w^{\mathcal{R}}$ is not regular

- Let $C = \{s \in \Sigma^* \;\; \exists w \in \Sigma.\ s = w \cdot w^{\mathcal{R}}$.
- Let p be a proposed pumping lemma constant for language C.
- Let $w = \mathsf{a}^p\mathsf{bba}^p$.
  - $w = w^{\mathcal{R}}$, thus $w \in C$.
  - $|w| = 2p + 1 > p$. Thus w satisfies the conditions of the pumping lemma.
- Let x, y, and z be any strings with $w = xyz$, $|y| = k > 0$ and $|xy| \leq p$.
  - Note that $x \in \mathsf{a}^*$ and $y \in \mathsf{a}^+$.
  - Therefore, $xy^0z = \mathsf{a}^{p-k}\mathsf{ba}^p \notin C$ becauls $p - k \neq p$.
- Thus, it is not possible to choose strings x, y and z that satisfy the conditions of the pumping lemma.
- Therefore, C is not regular.

The Pumping Game

- We can see this as a game between You and an Adversary. You want to show that language A is not regular, and the adversary want to thwart you.
- The Adversary has to make the first move by stating a value for p.
- Based on the value for p, You put forward a string $w \in A$.
- The Adversary now gives strings x, y, and z such that $w = xyz$, $|y| > 0$, and $|xy| \leq p$.
- If You can find a value for i such that $xy^iz \notin A$, then You win – you've shown that the language is not regular. ☺
  Otherwise, the Adversary wins. ☹

One More Example

- Let $A = \mathtt{1}^p$ where p is a prime number.
- Is A regular?

One More Example

- Let $A = 1^p$ where p is a prime number.
- A is not regular.
- Proof (by the pumping lemma, of course):
  - The Adversary proposes n, the pumping lemma constant for A.
  - You choose a prime q with $q > n$.
    Thus, $1^q \in A$. Give the Adversary the string 1^q.
  - The Adversary breaks 1^q into strings x, y and z such that $xyz = 1^q$, and $|y| > 0$.
  - You choose $i = (1 + q)$.
    The string $xy^{(1+q)}z$ has length

$$\begin{aligned} |x| + (1+q)|y| + |z| &= (|x| + |y| + |z|) + q|y| \\ &= q + q|y| \\ &= q(1 + |y|) \end{aligned}$$

    which is not prime (because $|y| > 0$).
  - Thus, $xy^{(1+q)}z \notin C$.
  - Therefore, You win; the Adversary loses; C is not regular. ☺

A Few Remarks

- WARNING: There are non-regular languages that satisfy the pumping lemma. For example,

$$\Sigma = \{\mathsf{a}, \mathsf{b}, \mathsf{c}\}$$
$$A = (\mathsf{aa}^*\mathsf{c})^n(\mathsf{bb}^*\mathsf{c})^n \cup \Sigma^*\mathsf{cc}\Sigma^*$$

  The language A is not regular, but it satisfies the conditions of the pumping lemma.

- Satisfying the conditions of the pumping lemma is a necessary but not sufficient condition for showing that a language is regular.

- If A is finite (i.e. $|A|$ is finite), then A trivially satisfies the pumping lemma. Let

$$p = 1 + \max_{w \in A} |w|$$

  There are no strings in A with length at least p, and the conditions of the pumping lemma are (vacuously) satisfied.